Statistik für Ingenieure
5 Schließende Statistik

Prof. Dr. Hans-Jörg Starkloff

TU Bergakademie Freiberg
Institut für Stochastik

Wintersemester 2019/2020
letzte Änderung: 6.1.2020

5 Schließende Statistik

5.1 Statistische Tests (Signifikanztests)

- Mit Hilfe von statistischen Tests (Signifikanztests) überprüft man, ob die vorhandenen Daten mit bestimmten Annahmen an die Verteilung der entsprechenden Zufallsgrößen im stochastischen Modell verträglich sind.
- Dabei muss man berücksichtigen, dass bedingt durch die Zufallssituation und die zufällige Streuung der Realisierungen der Zufallsgrößen im Allgemeinen keine 100%-ig richtigen Entscheidungen (die Annahmen an die Verteilung stimmen / stimmen nicht) getroffen werden können, sondern dass jede Entscheidung auch fehlerhaft sein kann.
- Deshalb versucht man die Wahrscheinlichkeit dafür, dass die Entscheidung fehlerhaft ist, zu kontrollieren.

Beispielaufgabe: Waschmittelpackungen

- Bei einem Verbrauchertest für Waschmittel werde auch die Abfüllmenge kontrolliert. Dabei ergaben sich bei 10 zufällig ausgewählten 5 kg Packungen einer bestimmten Sorte folgende Abfüllmengen (in kg) :

  $$4.6\ ,\ 4.95\ ,\ 4.8\ ,\ 4.9\ ,\ 4.75\ ,\ 5.05\ ,\ 4.9\ ,\ 5.1\ ,\ 4.85\ ,\ 4.95\,.$$

  Ist auf der Basis dieser Beobachtungswerte die Auffassung vertretbar, dass die Packungen im Mittel weniger Waschmittel als angegeben enthalten ?
- Wir modellieren die tatsächliche Abfüllmenge (in kg) einer Waschmittelpackung als Zufallsgröße X .
- Berechnete Schätzwerte für den Erwartungswert, die Standardabweichung und die Varianz der Merkmalsgröße sind:

  $$\overline{x} = 4.885\,, \qquad s = 0.145\,, \qquad s^2 = 0.0211\,.$$

Überlegungen zur Beispielaufgabe

- Der Erwartungswert μ ist unbekannt.
- Zu überprüfen ist die Richtigkeit der Vermutung, dass der Erwartungswert μ kleiner ist als der Sollwert $\mu_0 = 5$.
- Dies kann aber nicht einfach aus der Tatsache

$$\overline{x} = 4.885 < 5 = \mu_0$$

  gefolgert werden.
- Man kann schließlich zufällig eine Stichprobe mit geringen Abfüllmengen erwischt haben.

Hintergrund des Tests für die vereinfachte Beispielaufgabe

- **Annahme:** $X \sim \mathrm{N}(\mu_0, \sigma^2) = \mathrm{N}(5, 0.025)$.
- Folglich gilt: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu_0, \frac{\sigma^2}{n}\right) = \mathrm{N}\left(5, \frac{0.025}{n}\right)$,

  falls $X_i \sim \mathrm{N}(\mu_0, \sigma^2), i = 1, \ldots, n$ i.i.d. (mathem. Stichprobe).
- Insbesondere: $\overline{X} \sim \mathrm{N}(5, 0.0025)$ und $T := \frac{\overline{X}-5}{0.05} \sim \mathrm{N}(0,1)$.
- Für die konkrete Stichprobe gilt: $t = -2.3$.
- Realisierungen t der Testgröße T mit $t < z_{0.05} = -1.645$ (0.05-Quantil der Standardnormalverteilung) sind sehr selten, deshalb geht man bei einer auftretenden Realisierung der Testgröße in diesem Bereich (wie hier im Beispiel) eher davon aus, dass die gemachte Annahme („der wahre Erwartungswert von X ist 5“) falsch ist, die Abweichungen vom Sollwert also „signifikant“ („statistisch gesichert“) sind.

Grundlegende Überlegungen zu statistischen Tests

- **Aufstellen der Hypothesen:**

  Man formuliert 2 Hypothesen, die Nullhypothese H_0 und die Alternativhypothese H_A (oft auch mit H_1 bezeichnet)

  z.B. $H_0 : \mu = \mu_0$ und $H_A : \mu \neq \mu_0$
  oder $H_0 : \mu = \mu_0$ und $H_A : \mu < \mu_0$.

  Beachte: Die Hypothese, die statistisch abgesichert werden soll, sollte als Alternativhypothese formuliert werden!

- **2 mögliche Entscheidungen beim Testen:**
  1. H_0 wird verworfen: Es gibt in der erhobenen Stichprobe starke Hinweise darauf, dass H_0 nicht gelten kann, also H_A gelten muss. Diese Hinweise sind so stark, dass man nicht von einem zufälligen Zustandekommen ausgehen kann.
  2. H_0 wird nicht verworfen: Man hat keine Hinweise gefunden, die gegen H_0 sprechen. Alle aufgetretenen Effekte könnten genausogut zufallsbedingt sein.

Grundlegende Überlegungen zu statistischen Tests

- **Statistisches Testproblem:** Aufgabenstellung zwischen der Gültigkeit von H_0 und H_A zu unterscheiden.
- **Statistischer Test:** formale Entscheidungsregel für eine der zwei Möglichkeiten.
- **Mögliche Fehler beim Testen:**
  - Fehler 1. Art: man verwirft H_0, obwohl H_0 richtig ist.
  - Fehler 2. Art: man verwirft H_0 nicht, obwohl H_0 falsch ist.

$\Rightarrow$ Tests sind so zu konstruieren, dass beide Fehler möglichst klein sind.

- **Aber** es können nicht beide Fehler gleichzeitig kontrolliert werden.

$\Rightarrow$ Man gibt sich eine (relativ kleine) obere Schranke für die Wahrscheinlichkeit des Fehlers 1. Art vor, die nicht überschritten werden soll – das sogenannte Signifikanzniveau α.

- Übliche Werte für das Signifikanzniveau α sind 0.05 oder 0.01.

Grundlegende Überlegungen zu statistischen Tests

- In der Regel wird ein statistischer Test so konstruiert, dass er unter allen Tests, für die die Wahrscheinlichkeit des Fehlers 1. Art das gegebene Signifikanzniveau nicht überschreitet, den Fehler 2. Art minimiert.
- Wie erhält man eine Entscheidungsregel für ein gegebenes Testproblem?
- Im obigen Beispiel würde man intuitiv so vorgehen:
  - Liegt die Schätzung $\overline{x}$ für μ über oder nur knapp unter $\mu_0 = 5$, so kann man nicht mit hinreichender Sicherheit schließen, dass $H_0 : \mu \geq \mu_0 = 5$ nicht gilt.
  - Liegt hingegen $\overline{x}$ unter einem kritischen Wert deutlich unter $\mu_0 = 5$, so kann man die Nullhypothese verwerfen.
  - Wie weit der kritische Wert unter μ_0 liegen muss, hängt vom Signifikanzniveau α und dem Stichprobenumfang ab (und von der unbekannten Varianz).

Allgemeine Struktur der Entscheidungsregel

- Im Allgemeinen besteht die Entscheidungsregel für ein Testproblem aus einer Testgröße T und einem kritischen Bereich K_α.
- Testgröße T:
  - ist eine Stichprobenfunktion (d.h. eine Funktion der mathematischen Stichprobe $X_1, ..., X_n$), also eine Zufallsgröße;
  - ist bei Parametertests oft eine Schätzfunktion für den zu testenden Parameter oder davon abgeleitet (im Beispiel $\overline{X}$);
  - hat eine bekannte Verteilung bei Gültigkeit der Nullhypothese.
  - Setzt man statt der mathematischen Stichprobe eine konkrete Stichprobe $x_1, ..., x_n$ ein, so erhält man eine reelle Zahl t als Realisierung der Zufallsgröße T.
- Kritischer Bereich (Ablehnungsbereich) K_α:
  - ist von α abhängig;
  - wird so konstruiert, dass $\mathbf{P}(T \in K_\alpha | H_0) \leq \alpha$ gilt.
  - Im Beispiel ist $K_\alpha = \{t \in \mathbb{R} : t < t_\alpha\}$, wobei t_α der oben erwähnte kritische Wert ist.

Entscheidung beim Test

- Die Entscheidung lautet dann: ist $t \in K_\alpha$, so wird H_0 verworfen, andernfalls nicht.
- Alternative Entscheidungsregel (zumeist in statistischer Software umgesetzt):
  - Berechnung eines *p*-Werts (*p*-value): $p = \min\{\alpha \,:\, t \in K_\alpha\}$;
  - H_0 wird verworfen, wenn $p \leq \alpha$, bei $p > \alpha$ wird H_0 beibehalten.

Allgemeiner Testablauf

Allgemeiner Ablauf eines statistischen Tests:

1. **Aufstellen der Hypothesen**
2. **Festlegen des Signifikanzniveaus** α
3. **Bestimmen der Testgröße** T
4. **Berechnung der Realisierung** t **der Testgröße** T auf der Basis der konkreten Stichprobe $(x_1, \ldots, x_n)$
5. **Bestimmen des kritischen Bereichs** K_α **bzw. des** p**-Wertes**
6. **Testentscheidung:**

   $t \in K_\alpha \quad \Leftrightarrow \quad p \leq \alpha \quad \Rightarrow$ Ablehnung von H_0 ;
   $t \notin K_\alpha \quad \Leftrightarrow \quad p > \alpha \quad \Rightarrow$ Stichprobe spricht nicht gegen H_0 .

7. **Schlussfolgerung für die gegebene Aufgabenstellung**
   Im Beispiel :
   Die Vermutung, dass das Gewicht der Waschmittelpackungen systematisch geringer als 5 kg ist, ist (nicht) statistisch abgesichert.

Interpretation der Testergebnisse

- Beim Testen wird nur die Wahrscheinlichkeit für den Fehler 1. Art kontrolliert, d.h. $\mathbf{P}(H_0 \text{ ablehnen } \mid H_0 \text{ ist wahr}) \leq \alpha$.
- Wenn also H_0 tatsächlich gilt, wird man sich nur in $\alpha \cdot 100\%$ der Fälle für H_A entscheiden.
- Die Entscheidung für H_A ist in diesem Sinn **statistisch abgesichert**.
- Bei einer Entscheidung gegen H_0 und damit für H_A spricht man von einem signifikanten Ergebnis.
- Die Wahrscheinlichkeit für den Fehler 2. Art wird nicht kontrolliert.

$\Rightarrow$ Eine Entscheidung H_0 beizubehalten ist nicht statistisch abgesichert.

$\Rightarrow$ Kann man H_0 nicht verwerfen, bedeutet das daher nicht, dass man sich „aktiv“ für H_0 entscheidet, es spricht nur nichts gegen H_0.

Auswahl eines geeigneten Tests

- Da es eine Vielzahl unterschiedlicher Tests gibt, ist die Auswahl eines geeigneten Tests eine wichtige Aufgabe.
- Bei dieser Auswahl spielen unter anderem eine Rolle
  - das Skalenniveau des Merkmals oder der Merkmale;
  - die Stichprobensituation: eine Stichprobe von reellen Werten / eine vektorielle Stichprobe (eine gepaarte oder verbundene Stichprobe) / zwei (unabhängige) Stichproben / mehr als zwei (unabhängige) Stichproben;
  - Vorkenntnisse (z.B. durch vorangegangene Tests) oder Annahmen an die Verteilung der Merkmalszufallsgröße(n);
  - die zu lösende Aufgabenstellung, z.B. im Hinblick auf mögliche unterschiedliche Alternativhypothesen zu einer gewählten Nullhypothese.
- Viele Tests sind in Statistikcomputerprogrammen verfügbar, auch in R. Dann ist neben der Auswahl eines geeigneten Tests auch wichtig, die Vorgehensweise bzw. den Aufruf zu kennen und die Ergebnisse richtig auszuwerten.

Überblick: Tests für Merkmale mit stetiger Skala

Tests für Merkmale mit stetiger Skala

Problemstellung		*Voraussetzung an die Verteilung*	Stichprobensituation: Eine Stichprobe X	Gepaarte Stichproben D = X - Y	Zwei Stichproben X, Y	Mehrere Stichproben $X_1, \ldots, X_k$, k=3,4,...
Lageparameter		*normal*	Ein-Stichproben-t-Test	Ein-Stichproben-t-Test für D bzw. Gepaarter t-Test für X und Y	*Streuungen von X und Y unbekannt und gleich:* Zwei-Stichproben-t-Test *Streuungen von X und Y unbekannt und nicht gleich:* WELCHs-t-Test	*Streuungen von $X_1, \ldots, X_k$ unbekannt und gleich:* Varianzanalyse (ANOVA)
Lageparameter		*stetig*	Vorzeichentest *symmetrische Verteilung:* WILCOXON-Vorzeichen-Rang-Test	Vorzeichentest für D *symmetrische Verteilung:* WILCOXON-Vorzeichen-Rang-Test	WILCOXON-Rang-Summen-Test	KRUSKAL-WALLIS-Test
Streuungsparameter		*normal*	χ^2-Test auf Streuung	χ^2-Test auf Streuung für D	F-Test	BARTLETT-Test
Streuungsparameter		*stetig*	?	?	FLIGNER-Test	FLIGNER-Test
Verteilung	normal		SHAPIRO-WILK-Test	SHAPIRO-WILK-Test für D	2 SHAPIRO-WILK-Tests für X und Y	*k* SHAPIRO-WILK-Tests für $X_1, \ldots, X_k$
Verteilung	identisch	*stetig*	X	X	Zwei-Stichproben-KOLMOGOROFF-SMIRNOW-Test	?
Verteilung	konkret		χ^2-Anpassungs-Test	χ^2-Anpassungs-Test für D	2 χ^2-Anpassungs-Tests für X und Y	*k* χ^2-Anpassungs-Tests für $X_1, \ldots, X_k$
Verteilung	konkret	*stetig*	Ein-Stichproben-KOLMOGOROFF-SMIRNOW-Test	Ein-Stichproben-KOLMOGOROFF-SMIRNOW-Test für D	2 Ein-Stichproben-KOLMOGOROFF-SMIRNOW-Tests für X und Y	*k* Ein-Stichproben-KOLMOGOROFF-SMIRNOW-Tests für $X_1, \ldots, X_k$

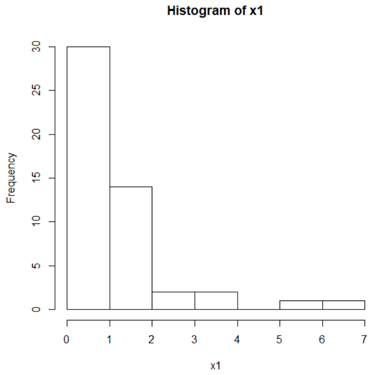
5.1.1 Tests für eine Stichprobe mit stetiger Skala

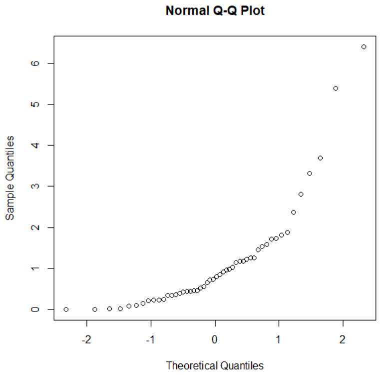
a) Shapiro-Wilk-Test

- Mit dem Shapiro-Wilk-Test überprüft man, ob die Daten mit einer Normalverteilung verträglich sind.
- **Geg.:** konkrete Stichprobe $x_1, \ldots, x_n$.
- **Vor.:** Merkmalszufallsgröße X auf stetiger Skala; repräsentative Stichprobe.
- **Hyp.:** H_0 : X ist normalverteilt ; H_A : X ist nicht normalverteilt
- **R-Aufruf:** `shapiro.test()`
- **Bem.:**
  - Die Parameter der vermuteten Normalverteilung (Erwartungswert und Varianz) müssen nicht bekannt sein.
  - Der Test reagiert sensibel auf Ausreißer.
  - Der Test ist relativ anfällig gegenüber Bindungen, deshalb sollten die Werte nicht stark gerundet sein.
  - Die Teststärke ist insbesondere bei kleinen Stichprobenumfängen größer als bei allgemeinen Anpassungstests, wie dem Kolmogorow-Smirnow-Test oder dem χ^2-Anpassungstest.

Bsp. Shapiro-Wilk-Test für exponentialverteilte Daten

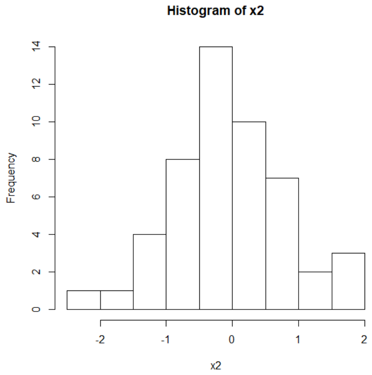
```
> x1=rexp(50)                   # Simulation der exponentialverteilten Werte
> shapiro.test(x1)
    Shapiro-Wilk normality test
data: x1
W = 0.87719, p-value = 9.203e-05              # W ist Wert der Teststatistik
> hist(x1)                                                     # Histogramm
> qqnorm(x1)                             # Q-Q-Plot bzgl. Normalverteilung
```

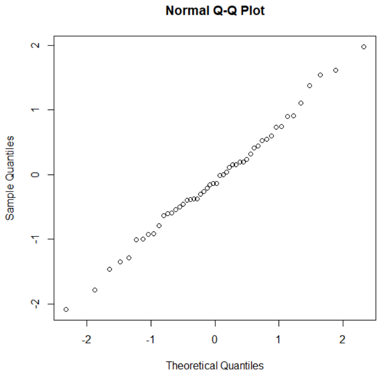




Bsp. Shapiro-Wilk-Test für normalverteilte Daten

```
> x2=rnorm(50)                         # Simulation der normalverteilten Werte
> shapiro.test(x2)
    Shapiro-Wilk normality test
data: x2
W = 0.99268, p-value = 0.9885                  # W ist Wert der Teststatistik
> hist(x2)                                                     # Histogramm
> qqnorm(x2)                           # Q-Q-Plot bzgl. Normalverteilung
```





b) Kolmogorow-Smirnow-Test

- Mit dem Kolmogorow-Smirnow-Test überprüft man, ob die Daten mit einer vorgebenen Verteilung verträglich sind.
- **Geg.:** konkrete Stichprobe $x_1, \ldots, x_n$.
- **Vor.:** Merkmalszufallsgröße X auf stetiger Skala; repräsentative Stichprobe.
- **Hyp.:**
  $H_0: \; F_X = F_0$ (Verteilungsfunktion von X ist F_0);
  $H_A: \; F_X \neq F_0$ (Verteilungsfunktion von X ist nicht F_0).
- **R-Aufruf:** `ks.test(,)`
- **Bem.:**
  - Die Verteilungsfunktion F_0 muss vollständig bekannt sein, insbesondere alle Parameter.
  - Es gibt Varianten des Tests für spezielle Fälle mit geschätzten Parametern.
  - Der Test ist relativ anfällig gegenüber Bindungen, deshalb sollten die Werte nicht stark gerundet sein.

Bsp. Kolmogorow-Smirnow-Test mit R

```
> x1=rexp(50)        # Simulation der exponentialverteilten Werte (Parameter=1)
> ks.test(x1,"pexp")           # Test auf Exponentialverteilung mit Parameter=1

    One-sample Kolmogorov-Smirnov test

data: x1
D = 0.086285, p-value = 0.8196                  # D ist Wert der Teststatistik
alternative hypothesis: two-sided


> ks.test(x1,"pexp",2)         # Test auf Exponentialverteilung mit Parameter=2

    One-sample Kolmogorov-Smirnov test

data: x1
D = 0.3334, p-value = 1.862e-05                 # D ist Wert der Teststatistik
alternative hypothesis: two-sided


> x2=rnorm(50)                         # Simulation der normalverteilten Werte
> ks.test(x2,"pnorm")

One-sample Kolmogorov-Smirnov test

data: x2
D = 0.1081, p-value = 0.566                     # D ist Wert der Teststatistik
alternative hypothesis: two-sided
```

c) χ^2- Anpassungstest

- Mit dem χ^2- Anpassungstest überprüft man, ob die Daten mit einer vorgebenen Verteilung verträglich sind.

- **Geg.:** konkrete Stichprobe $x_1, \ldots, x_n$.

- **Vor.:** Merkmalszufallsgröße X auf stetiger Skala (auch für andere möglich); repräsentative Stichprobe.

- **Hyp.:**
  $H_0: \; F_X = F_0$ (Verteilungsfunktion von X ist F_0) ;
  $H_A: \; F_X \neq F_0$ (Verteilungsfunktion von X ist nicht F_0) .

- **R-Aufruf:** `chisq.test(,)`

Bemerkungen zum χ^2- Anpassungstest

- Der χ^2-Anpassungstest für stetige Daten basiert auf einer Klasseneinteilung der Stichprobe und dem Vergleich der theoretischen Häufigkeiten der Werte in den Klassen mit den empirischen Häufigkeiten.
- Die Testgröße ist unter H_0 asymptotisch χ^2-verteilt, dies ist eine häufiger vorkommende statistische Prüfverteilung mit einem Parameter, der Anzahl der Freiheitsgrade genannt wird. Sie kann nur nichtnegative Werte annehmen.
- Die theoretische Häufigkeit sollte pro Klasse mindestens 5 sein.
- Der Wert der Testgröße (und damit ggf. das Testergebnis) hängt von der gewählten Klasseneinteilung ab, außerdem ist es nur ein asymptotischer Test.

Bsp. χ^2- Anpassungstest mit R

```
> x2=rnorm(50)                        # Simulation der normalverteilten Werte
> x2_cut=cut(x2,breaks=c(-3,-2,-1,0,1,2))                  # Klasseneinteilung
> table(x2_cut)
x2_cut
(-3,-2] (-2,-1] (-1,0] (0,1] (1,2]
1 5 22 17 5
> freq_emp=vector()                    # Vektor der empirischen Häufigkeiten
> for(i in 1:5) freq_emp[i]=table(x2_cut)[[i]]
> freq_emp
[1] 1 5 22 17 5
> freq_th=c(pnorm(-2),pnorm(-1)-pnorm(-2), pnorm(0)-pnorm(-1),
+ pnorm(1)-pnorm(0),1-pnorm(1))
> freq_th                             # Vektor der theoretischen Häufigkeiten
[1] 0.02140023 0.13590512 0.34134475 0.34134475 0.13590512
> chisq.test(freq_emp,freq_th)

Pearson's Chi-squared test

data: freq_emp and freq_th
X-squared = 3.0011, df = 4, p-value = 0.5576

Warning message:
In chisq.test(freq_emp, p=freq_th) :
Chi-Quadrat-Approximation kann inkorrekt sein
```

d) Ein-Stichproben-t-Test

- Mit dem Ein-Stichproben-t-Test werden Annahmen über den Erwartungswert einer normalverteilten Grundgesamtheit bei unbekannter Varianz überprüft.
- **Geg.:** konkrete Stichprobe $x_1, \ldots, x_n$.
- **Vor.:** normalverteilte Merkmalszufallsgröße X mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 ; repräsentative Stichprobe.
- **Hyp.:**
  $H_0: \ \mu = \mu_0$ (μ_0 ist eine gegebene (Soll-)Größe);
  $H_A: \ \mu \neq \mu_0$ (zweiseitig) bzw. $\mu < \mu_0$ oder $\mu > \mu_0$ (einseitig).
- **R-Aufruf:** `t.test()`
- **Bem.:** Die Testgröße ist hier $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$, diese ist unter H_0 $t-$verteilt mit $n-1$ Freiheitsgraden. Die $t-$Verteilung oder Student-Verteilung ist eine weitere oft genutzte statistische Prüfverteilung mit einem Parameter („Anzahl der Freiheitsgrade“).

Bsp. Ein-Stichproben-t-Test mit R

- Simulation von Realisierungen $\mathrm{N}(0,1)$-verteilter Zufallsgrößen.

```
x=rnorm(50)
```

- Zweiseitiger $t-$Test für $H_0 : \mu = 0\,,\ H_A : \mu \neq 0:$

```
> t.test(x)

One Sample t-test

data: x
t = -0.63253, df = 49, p-value = 0.53
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.3206896 0.1671406
sample estimates:
mean of x
-0.07677448
```

Bsp. Ein-Stichproben-t-Test (einseitig) mit R

- Einseitiger $t-$Test für $H_0 : \mu = 0,\ H_A : \mu < 0$:

```
> t.test(x,alternative="less")

One Sample t-test

data: x
t = -0.63253, df = 49, p-value = 0.265
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
-Inf 0.1267193
sample estimates:
mean of x
-0.07677448
```

- Einseitiger $t-$Test für $H_0 : \mu = 0\,,\ H_A : \mu > 0$:

```
> t.test(x,alternative="greater")

One Sample t-test

data: x
t = -0.63253, df = 49, p-value = 0.735
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
-0.2802683 Inf
sample estimates:
mean of x
-0.07677448
```

Bsp. Ein-Stichproben-t-Test mit R Fortsetzung

- Zweiseitiger $t-$Test für $H_0 : \mu = 1\,,\; H_A : \mu \neq 1 :$

```
> t.test(x,mu=1)

One Sample t-test

data: x
t = -8.8714, df = 49, p-value = 9.175e-12
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
-0.3206896 0.1671406
sample estimates:
mean of x
-0.07677448
```

- Zweiseitiger $t-$Test für $H_0 : \mu = -0.1\,,\; H_A : \mu \neq -0.1 :$

```
> t.test(x,mu=-0.1)

One Sample t-test

data: x
t = 0.19135, df = 49, p-value = 0.849
alternative hypothesis: true mean is not equal to -0.1
95 percent confidence interval:
-0.3206896 0.1671406
sample estimates:
mean of x
-0.07677448
```

e) χ^2-Test auf Streuung

- Mit dem χ^2-Test auf Streuung werden Annahmen über die Varianz einer normalverteilten Grundgesamtheit bei unbekanntem Erwartungswert überprüft.
- **Geg.:** konkrete Stichprobe $x_1, \ldots, x_n$.
- **Vor.:** normalverteilte Merkmalszufallsgröße X mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 ; repräsentative Stichprobe.
- **Hyp.:**
  $H_0:\ \sigma^2 = \sigma_0^2$ (σ_0^2 ist eine gegebene (Soll-)Größe) ;
  $H_A:\ \sigma^2 \neq \sigma_0^2$ (zweiseitig) bzw. $\sigma^2 < \sigma_0^2$ oder $\sigma^2 > \sigma_0^2$ (einseitig) .
- **R-Aufruf:** `sigma.test()` aus Zusatzpaket „TeachingDemos".
- Die **Testgröße** ist hier $T = \dfrac{(n-1)S^2}{\sigma_0^2}$, diese ist unter H_0 χ^2-verteilt mit $n-1$ Freiheitsgraden.

Bsp. 1 zweiseitiger χ^2-Test auf Streuung mit R

Voraussetzung ist, dass das Programmpaket „TeachingDemos“ vorher installiert wurde.

```
> require(TeachingDemos)                               # Laden des Programmpakets
> x=rnorm(50)                            # Simulation der normalverteilten Werte
> sigma.test(x)

One sample Chi-squared test for variance

data: x
X-squared = 36.094, df = 49, p-value = 0.1704
alternative hypothesis: true variance is not equal to 1
95 percent confidence interval:
0.5139954 1.1438471
sample estimates:
var of x
0.7366122
```

Bsp. 2 einseitiger χ^2-Test auf Streuung mit R

```
> require(TeachingDemos)                             # Laden des Programmpakets
> x=rnorm(50)                            # Simulation der normalverteilten Werte
> sigma.test(x,sigmasq=0.5,alternative="greater")

One sample Chi-squared test for variance

data: x
X-squared = 72.188, df = 49, p-value = 0.01721
alternative hypothesis: true variance is greater than 0.5
95 percent confidence interval:
0.544087 Inf
sample estimates:
var of x
0.7366122
```

f) Vorzeichentest

- Der Vorzeichentest oder Zeichentest dient als Test über den Median einer stetigen Verteilung.
- **Geg.:** konkrete Stichprobe $x_1, \ldots, x_n$.
- **Vor.:** Merkmalszufallsgröße X auf stetiger Skala; repräsentative Stichprobe.
- **Hyp.:**
  $H_0: \ X_{0.5} = m$ (m ist ein vorgebener Wert für den Median);
  $H_A: \ X_{0.5} \neq m$.
- **R-Aufruf:** `binom.test(table(x<m))` (für Datenvektor `x`).
- Die Testgröße ist die Anzahl der Stichprobenwerte, die **größer oder gleich** dem hypothetischen Wert m für den Median sind. Sie ist unter H_0 binomialverteilt mit den Parametern n und $p = 0.5$. Der Test heißt deshalb auch Binomialtest (bzw. ist ein Spezialfall davon).

Bsp. Vorzeichentest

- Der Vorzeichentest wird auf simulierte exponentialverteilte mit Parameter $\lambda = 1$ Daten angewandt.

  Der theoretische Median einer solchen exponentialverteilten Zufallsgröße ist $x_{0.5} = \ln(2) = 0.6931472$.

-
```
>x=rexp(30)                  # Simulation der exponentialverteilten Werte
> binom.test(table(x<log(2)))

    Exact binomial test

data: table(x < log(2))
number of successes = 16, number of trials = 30, p-value = 0.8555
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.3432552 0.7165819
sample estimates:
probability of success
    0.5333333
```

Bsp. Vorzeichentest Fortsetzung

- Bei einem Test auf den (falschen) hypothetischen Medianwert $m = 1$ erhält man für diese Stichprobe folgenden Ausdruck.

- ```
> binom.test(table(x<1))

	Exact binomial test

data: table(x < 1)
number of successes = 9, number of trials = 30, p-value = 0.04277
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1473452 0.4939590
sample estimates:
probability of success
    0.3
```
```

Bsp. Vorzeichentest Erläuterung zur Fortsetzung

- Zur Erläuterung der R-Befehle seien hier die Stichprobe und Zwischenergebnisse mit angegeben.
-
```
> x
[1] 0.12452168 0.45299701 0.02058257 0.75440725 0.86050930
[6] 2.97866055 0.03318594 0.63691576 0.81718036 0.45254250
[11] 0.20732538 0.93757553 0.92931209 2.21512245 0.86975410
[16] 0.60563118 0.41212784 0.05024501 1.91634500 1.05197948
[21] 0.67901945 1.61321168 0.65232898 1.67017803 0.06047516
[26] 0.80740846 2.01478421 1.14940138 1.15195415 0.18380546
> x<1
[1] TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE
[11] TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE FALSE FALSE
[21] TRUE FALSE TRUE FALSE TRUE TRUE FALSE FALSE FALSE TRUE
> table(x<1)

 FALSE   TRUE
     9     21
```
- Die Erfolgsanzahl im Test (hier 9, die erste der durch `table(x<1)` zurückgegebene Zahl) ist also die Anzahl der Stichprobenwerte, für die die Bedingung (hier $x < 1$) **nicht** erfüllt ist.

Bsp. Vorzeichentest (einseitig)

- Einseitige Tests können auch durchgeführt werden.
-
```
> binom.test(table(x<1),alternative="less")
    Exact binomial test
data: table(x < 1)
number of successes = 9, number of trials = 30, p-value = 0.02139
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.4650727
sample estimates:
probability of success
    0.3
```

- Hier wird zum Niveau 0.05 die Hypothese $H_0 : \mathbf{P}(X \geq 1) = 0.5$ abgelehnt und die Alternative $H_A : \mathbf{P}(X \geq 1) < 0.5$ angenommen. Dies bedeutet auch für den Median, dass er signifikant kleiner als 1 ist.

g) WILCOXON-Vorzeichen-Rang-Test

- Beim WILCOXON-Vorzeichen-Rang-Test werden Hypothesen über das Symmetriezentrum (und damit den Median) einer stetigen Verteilung geprüft.
- **Geg.:** konkrete Stichprobe $x_1, \ldots, x_n$.
- **Vor.:** Merkmalszufallsgröße X mit stetiger und **symmetrischer** Verteilung ; repräsentative Stichprobe .
- **Hyp.:**
  H_0 : $X_{0.5} = m$ (m ist ein vorgebener Wert für den Median);
  H_A : $X_{0.5} \neq m$.
- **R-Aufruf:** `wilcox.test()` .
- Die Testgröße nutzt Rangzahlen der Werte $x_i - m$, $i = 1, \ldots, n$, und damit mehr Informationen als der Vorzeichentest.
- Bindungen können problematisch sein.

Bsp. Wilcoxon-Vorzeichen-Rang-Test

- Der Vorzeichentest wird auf simulierte $t-$verteilte (mit 10 Freiheitsgraden) Daten angewandt. Dies ist eine symmetrische stetige Verteilung mit dem theoretischen Median $x_{0.5} = 0$.
-
```
>x=rt(n=50,df=10)                  # Simulation von 50 t-verteilten Werten
> wilcox.test(x)

    Wilcoxon signed rank test with continuity correction

data: x
V = 800, p-value = 0.1179                                          # Annahme
alternative hypothesis: true location is not equal to 0
```
- Ein Test auf den (falschen) Median $m = 1$ ergibt:

```
> wilcox.test(x,mu=1)

    Wilcoxon signed rank test with continuity correction

data: x
V = 195, p-value = 1.983e-05                                     # Ablehnung
alternative hypothesis: true location is not equal to 1
```

5.1.2 Tests für eine gepaarte (verbundene) Stichprobe (stetige Skala)

- Gegeben sei eine konkrete Stichprobe $(x_i, y_i)\,, i = 1, \ldots, n\,,$ als Realisierungen von unabhängigen und identisch verteilten stetigen Zufallsvektoren $(X_i, Y_i)\,, i = 1, \ldots, n$. Für jedes i beziehen sich die Werte x_i und y_i auf ein und denselben Merkmalsträger, so dass die Zufallsgrößen X_i und Y_i nicht als unabhängig angesehen werden können.
- Macht die Differenzbildung $D_i = X_i - Y_i\,, i = 1, \ldots, n\,,$ inhaltlich Sinn, dann können die Tests aus 5.1.1. auf die neu berechnete Stichprobe $d_1, \ldots, d_n$ (die nun univariat ist) angewandt werden, man untersucht somit ein Einstichprobenproblem.
- Dabei sind insbesondere die Tests bezüglich der Lageparameter von Interesse, da dadurch eine eventuelle Verschiebung der Verteilung der Y_i zu den Größen X_i mit Hilfe eines Tests auf einen Median oder Erwartungswert 0 der Verteilung der Differenzzufallsgrößen $D_i\,, i = 1, \ldots, n$, überprüft werden kann.

a) Gepaarter $t-$Test

- Mit dem Ein-Stichproben-t-Test für $D = X - Y$ oder dem gepaarten $t-$Test für X und Y wird die Gleichheit der Erwartungswerte von X und Y bei einer normalverteilten Differenz $D = X - Y$ mit unbekannter Varianz überprüft.
- **Geg.:** konkrete gepaarte Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$.
- **Vor.:** normalverteilte Zufallsgröße $D = X - Y$ mit unbekannter Varianz σ^2; repräsentative Stichprobe.
- **Hyp.:**
  $H_0: \ \mathbf{E}X = \mathbf{E}Y$, $\quad H_A: \ \mathbf{E}X \neq \mathbf{E}Y$ (zweiseitiger Test) bzw.
  $H_A: \ \mathbf{E}X < \mathbf{E}Y$ oder $H_A: \ \mathbf{E}X > \mathbf{E}Y$ (einseitige Tests).
- **R-Aufruf:** `t.test(x,y,paired=TRUE)`
  bei Datenvektoren `x` und `y`.
- Ausreißer in den Daten können Probleme bereiten.

Bsp. 1 gepaarter $t-$Test

- Simulation einer gepaarten Stichprobe durch Beziehung: fester Wert 2 + simulierte normalverteilte zufällige Fehler für die $x-$ und $y-$Werte jeweils.

- ```
  > set.seed(123456)
  > x=2+rnorm(50,sd=0.1)
  > y=2+rnorm(50,sd=0.1)
  ```

- Berechnung der Differenzen und Shapiro-Wilk-Test auf Normalverteilung .

- ```
  > d=x-y
  > shapiro.test(d)
      Shapiro-Wilk normality test
  data: d
  W = 0.98946, p-value = 0.9328                                # Nichtablehnung
  ```

- Durchführung des Ein-Stichproben-t-Tests für d und des äquivalenten gepaarten $t-$Tests für x und y .

Bsp. 1 gepaarter $t-$Test Fortsetzung

- ```
  > t.test(d)
      One Sample t-test
  data: d
  t = 0.99744, df = 49, p-value = 0.3235                  # Nichtablehnung
  alternative hypothesis: true mean is not equal to 0
  95 percent confidence interval:
   -0.02040492 0.06062202
  sample estimates:
   mean of x
  0.02010855
  ```

- ```
  > t.test(x,y,paired=TRUE)
      Paired t-test
  data: x and y
  t = 0.99744, df = 49, p-value = 0.3235                  # Nichtablehnung
  alternative hypothesis: true difference in means is not equal to 0
  95 percent confidence interval:
   -0.02040492 0.06062202
  sample estimates:
  mean of the differences
      0.02010855
  ```

Bsp. 2 gepaarter $t-$Test

- Simulation einer gepaarten Stichprobe durch Beziehungen 2 (bei x) bzw. 3 (bei y) + simulierte normalverteilte zufällige Fehler.

- ```
  > set.seed(123456)
  > x=2+rnorm(50,sd=0.1)
  > y=3+rnorm(50,sd=0.05)
  ```

- Berechnung der Differenzen und Shapiro-Wilk-Test auf Normalverteilung .

- ```
  > d=x-y
  > shapiro.test(d)
      Shapiro-Wilk normality test
  data: d
  W = 0.97728, p-value = 0.4437                        # Nichtablehnung
  ```

- Durchführung des Ein-Stichproben-t-Tests für `d` und des äquivalenten gepaarten $t-$Tests für `x` und `y` .

Bsp. 2 gepaarter $t-$Test Fortsetzung

- ```
  > t.test(d)
      One Sample t-test
  data: d
  t = -60.197, df = 49, p-value < 2.2e-16                    # Ablehnung
  alternative hypothesis: true mean is not equal to 0
  95 percent confidence interval:
   -1.0169291 -0.9512261
  sample estimates:
  mean of x
  -0.9840776
  ```

- ```
  > t.test(x,y,paired=TRUE)
      Paired t-test
  data: x and y
  t = -60.197, df = 49, p-value < 2.2e-16                    # Ablehnung
  alternative hypothesis: true difference in means is not equal to 0
  95 percent confidence interval:
   -1.0169291 -0.9512261
  sample estimates:
  mean of the differences
      -0.9840776
  ```

b) Vorzeichentest für eine gepaarte Stichprobe

- Der Vorzeichentest für eine gepaarte Stichprobe ist ein Test über den Median 0 der stetigen Verteilung von $D = X - Y$. Bei Ablehnung der Nullhypothese kann man folglich auf eine unterschiedliche „mittlere Lage" der $x-$Werte und der $y-$Werte schließen.

- **Geg.:** konkrete gepaarte Stichprobe $(x_1, y_1) \ldots, (x_n, y_n)$.

- **Vor.:** Die Zufallsgröße $D = X - Y$ besitzt eine stetige Verteilung; es liegt eine repräsentative gepaarte Stichprobe vor.

- **Hypothesen:** $H_0: \ D_{0.5} = 0\,, \quad H_A: \ D_{0.5} \neq 0\,.$

- **R-Aufruf:** `binom.test(table(x<y))`
  bei Datenvektoren `x` und `y`.

- Bindungen können problematisch sein.

Bsp. Vorzeichentest für eine gepaarte Stichprobe

- Das Vorgehen ist analog zum 2. Anwendungsbeispiel für den gepaarten $t-$Test, jedoch mit exponentialverteilten Fehlern.
- ```
  > set.seed(123456)
  > x=2+rexp(50)                         # verschobene Exponentialverteilung
  > y=3+rexp(50)                         # verschobene Exponentialverteilung
  > shapiro.test(x-y)                            # Test auf Normalverteilung
      Shapiro-Wilk normality test
  data: x - y
  W = 0.91817, p-value = 0.002026                                # Ablehnung
  ```
- Vorzeichentest für eine gepaarte Stichprobe.
- ```
  > binom.test(table(x<y))

      Exact binomial test

  data: table(x < y)
  number of successes = 9, number of trials = 50, p-value = 5.614e-06
  alternative hypothesis: true probability of success is not equal to 0.5
  95 percent confidence interval:
   0.08576208 0.31436941
  sample estimates:
  probability of success
      0.18
  ```

c) Gepaarter Wilcoxon-Vorzeichen-Rang-Test

- Der gepaarte Wilcoxon-Vorzeichen-Rang-Test ist ein Test über das Symmetriezentrum 0 (und damit den Median 0) der stetigen Verteilung von $D = X - Y$. Bei Ablehnung der Nullhypothese kann man folglich auf eine unterschiedliche „mittlere Lage“ der $x-$ und der $y-$Werte schließen.
- **Geg.:** konkrete gepaarte Stichprobe $(x_1, y_1) \ldots, (x_n, y_n)$.
- **Vor.:** Die Zufallsgröße $D = X - Y$ besitzt eine stetige und **symmetrische** Verteilung; es liegt eine repräsentative gepaarte Stichprobe vor.
- **Hyp.:**
  H_0: Die Verteilung von $D = X - Y$ ist symmetrisch um 0;
  H_A: Die Verteilung von $D = X - Y$ ist symmetrisch um $c \neq 0$.
- **R-Aufruf:** `wilcox.test(x,y,paired=TRUE)`
  bei Datenvektoren `x` und `y`.
- Bindungen können problematisch sein.

Bsp. gepaarter Wilcoxon-Vorzeichen-Rang-Test

- Das Vorgehen ist analog zum 2. Anwendungsbeispiel für den gepaarten $t-$Test, jedoch werden hier $t-$verteilte Fehler verwendet.
- Simulation und Test auf Normalverteilung.
-
```
> set.seed(123456)
> x=2+0.1*rt(50,df=2)             # t-Verteilung mit 2 Freiheitsgraden
> y=3+0.1*rt(50,df=2)                  # t-Verteilung ist symmetrisch
> d=x-y
> shapiro.test(d)                         # Test auf Normalverteilung
    Shapiro-Wilk normality test
data: d
W = 0.94933, p-value = 0.03203                              # Ablehnung
```

Bsp. gepaarter WILCOXON-Vorzeichen-Rang-Test

- Anwendung des WILCOXON-Vorzeichen-Rang-Tests auf die Differenzen bzw. gepaart.
-
```
> wilcox.test(d)

    Wilcoxon signed rank test

data: d
V = 0, p-value = 7.79e-10                                  # Ablehnung
alternative hypothesis: true location is not equal to 0


> wilcox.test(x,y,paired=TRUE)

    Wilcoxon signed rank test

data: x and y
V = 0, p-value = 7.79e-10                                  # Ablehnung
alternative hypothesis: true location shift is not equal to 0
```

5.1.3 Tests für zwei oder mehr (unabhängige) Stichproben (stetige Skala)

- Von besonderer Bedeutung sind statistische Tests bezüglich der Lageparameter für die (unabhängigen) Zufallsgrößen X, Y bei zwei Stichproben bzw. $X_1, \ldots, X_k$ bei mehreren Stichproben.
- Um derartige Tests anwenden zu können, müssen im Allgemeinen vorher Annahmen über die Verteilungen der Einzelzufallsgrößen und teilweise auch über die Gleichheit der Varianzen überprüft werden.
- Es können wieder spezielle Tests verwendet werden, falls die Merkmalszufallsgrößen normalverteilt sind.
- Im Fall von nichtnormalverteilten Zufallsgrößen können oft rangbasierte (sogenannte verteilungsfreie) Tests verwendet werden. Diese können auch für normalverteilte Daten verwendet werden, sind dann aber nicht so effektiv wie die speziellen Tests.

a) Anpassungstests für mehrere Stichproben (stetige Skala)

- Statistische Tests über die Verteilung werden in dieser Situation oft so durchgeführt, dass für jede beteiligte reelle Stichprobe ein geeigneter Anpassungstest durchgeführt wird.
- So können beim Test auf Normalverteilung zwei (bzw. k) einzelne Shapiro-Wilk-Tests für X und Y (bzw. $X_1, \ldots, X_k$ im $k-$Stichprobenfall) durchgeführt werden.
- Analog können für andere Verteilungen zwei (bzw. k) einzelne χ^2-Anpassungstests oder Kolmogorow-Smirnow-Tests durchgeführt werden.
- Da bei der Durchführung mehrerer Tests, die nur zusammen eine Gesamtaussage erlauben, eine vorgegebene Wahrscheinlichkeit für einen Fehler 1. Art **für die Gesamtaussage** nicht mit dem entsprechenden Niveau der einzelnen beteiligten Tests übereinstimmt, sollte man in einer solchen Situation die sogenannte Bonferroni-Methode oder Bonferroni-Korrektur anwenden.

b) Bonferroni-Methode oder Bonferroni-Korrektur

- Angenommen eine Hypothese setzt sich aus k Einzelhypothesen wie folgt zusammen:

$$H_0 : H_0^1 \cap \ldots \cap H_0^k \,, \qquad H_A : H_A^1 \cup \ldots \cup H_A^k \,.$$

- Sind z.B. die k Zufallsgrößen $X_1, \ldots, X_k$ gegeben, erhält man

$$H_0 : \text{ alle } k \text{ ZG sind normalverteilt} \,,$$
$$H_A : \text{ mind. eine ZG } X_i \text{ ist nicht normalverteilt}$$

in obiger Weise aus den Einzelhypothesen

$$H_0^i : \; X_i \text{ ist normalverteilt} \,, \qquad H_A^i : \; X_i \text{ ist nicht normalverteilt} \,.$$

- Man führt nun k Tests bezüglich der Einzelhypothesen H^i durch, und entscheidet dann wie folgt:
  Man verwirft H_0, wenn mindestens ein Einzeltest die Nullhypothese H_0^i verwirft, sonst behält man H_0 bei.

Fortsetzung BONFERRONI-Korrektur

- Führt man die Einzeltests jeweils zum Signifikanzniveau $\tilde{\alpha}$ durch und bezeichne A_j, $j = 1, \ldots, k$, das zufällige Ereignis, dass der $j-$te Test seine Nullhypothese ablehnt, so gilt unter der Annahme der Unabhängigkeit der Ereignisse A_j und bei kleinem $\tilde{\alpha}$:

$$\begin{aligned}
\alpha &= \mathbf{P}(H_0 \text{ wird verworfen} \mid H_0 \text{ wahr}) \\
&= \mathbf{P}(A_1 \cup \ldots \cup A_k \mid H_0 \text{ wahr}) \\
&= 1 - \mathbf{P}({A_1}^{c} \cap \ldots \cap {A_k}^{c} \mid H_0 \text{ wahr}) \\
&= 1 - \mathbf{P}({A_1}^{c} \mid H_0 \text{ wahr}) \cdot \ldots \cdot \mathbf{P}({A_k}^{c} \mid H_0 \text{ wahr}) \\
&= 1 - (1-\tilde{\alpha})^k = 1 - 1 + k\tilde{\alpha} - \binom{k}{2}\tilde{\alpha}^2 + \ldots + (-1)^k \tilde{\alpha}^k \\
&\approx k\tilde{\alpha}\,.
\end{aligned}$$

- Folglich sollte man als Niveau der Einzeltests $\tilde{\alpha} = \dfrac{\alpha}{k}$ wählen.

c) $F-$Test für Varianzen zweier normalverteilter Merkmale

- Der $F-$Test dient zum Vergleich der Varianzen zweier unabhängiger normalverteilter Merkmale mit unbekannten Erwartungswerten.
- **Geg.:** 2 konkrete Stichproben $x_1, \ldots, x_n$ und $y_1, \ldots, y_m$ (die Stichprobenumfänge können unterschiedlich sein).
- **Vor.:** Die Zufallsgrößen X und Y sind unabhängig und normalverteilt mit (unbekannten) Erwartungswerten μ_X und μ_Y und Varianzen σ_X^2 und σ_Y^2; repräsentative Stichproben.
- **Hyp.:** $H_0: \ \sigma_X^2 = \sigma_Y^2, \quad H_A: \ \sigma_X^2 \neq \sigma_Y^2$ (zweiseitiger Test).
- **R-Aufruf:** `var.test(,)`.
- Ausreißer in den Daten können Probleme bereiten.
- Die Testgröße ist $F = \dfrac{S_X^2}{S_Y^2}$, sie ist unter H_0 $F-$verteilt mit $(n-1, m-1)$ Freiheitsgraden. Einseitige Tests sind auch möglich.

Bsp. $F-$Test für Varianzen zweier normalverteilter Merkmale ($\alpha = 0.05$)

- Simulation der Stichproben und Test auf Normalverteilung (mit Bonferroni-Korrektur).

-
```
> set.seed(123456)
> x=rnorm(30)
> y=rnorm(40)
> shapiro.test(x)

	Shapiro-Wilk normality test

data:  x
W = 0.97073, p-value = 0.5594               # >0.05/2, also Nichtablehnung

> shapiro.test(y)

	Shapiro-Wilk normality test

data:  y
W = 0.97027, p-value = 0.3671               # >0.05/2, also Nichtablehnung
```

Fortsetzung Bsp. $F-$Test

- Durchführung $F-$Test.
-
```
> var.test(x,y)

    F test to compare two variances

data: x and y
F = 0.94926, num df = 29, denom df = 39, p-value = 0.8952 # Nichtabl.
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4838555 1.9295306
sample estimates:
ratio of variances
    0.9492611
```

d) Bartlett-Test für Varianzen von Normalverteilungen

- Der Bartlett-Test dient zum Vergleich der Varianzen mehrerer unabhängiger normalverteilter Merkmale.
- **Geg.:** k Stichproben $x_{11}, \ldots, x_{1n_1}$ usw. bis $x_{k1}, \ldots, x_{kn_k}$ (die Stichprobenumfänge können unterschiedlich sein).
- **Vor:** Die Zufallsgrößen X_i , $i = 1, \ldots, k$, sind unabhängig und normalverteilt mit (unbekannten) Erwartungswerten μ_i und Varianzen σ_i^2 jeweils; repräsentative Stichproben.
- **Hyp.:** $H_0: \ \sigma_1^2 = \ldots = \sigma_k^2$, $\quad H_A: \ \sigma_i^2 \neq \sigma_j^2$ für mindestens ein Paar (i, j) .
- **R-Aufruf:** `bartlett.test()` .
- Ausreißer in den Daten können Probleme bereiten.
- Der Test ist ein asymptotischer Test, als Faustregel wird $n_i \geq 5$, $i = 1, \ldots, k$, empfohlen.
- Einseitige Tests sind hier nicht möglich.

Bsp. Bartlett-Test für Varianzen

- Simulation der Stichproben und Test auf Normalverteilung
- ```
  > set.seed(123456)
  > x1=rnorm(30)                                              # N(0,1)
  > x2=rnorm(30)                                              # N(0,1)
  > x3=rnorm(50,mean=1,sd=2)                                  # N(1,4)
  > shapiro.test(x1)

      Shapiro-Wilk normality test

  data: x1
  W = 0.97073, p-value = 0.5594              # >0.05/3, also Nichtablehnung

  > shapiro.test(x2)

      Shapiro-Wilk normality test

  data: x2
  W = 0.95518, p-value = 0.2321              # >0.05/3, also Nichtablehnung

  > shapiro.test(x3)

      Shapiro-Wilk normality test

  data: x3
  W = 0.97196, p-value = 0.2775              # >0.05/3, also Nichtablehnung
  ```
  ```

Fortsetzung Bsp. Bartlett-Test

- Durchführung Bartlett-Test.
- ```
  > bartlett.test(list(x1,x2,x3))

      Bartlett test of homogeneity of variances

  data: list(x1, x2, x3)
  Bartlett's K-squared = 19.967, df = 2, p-value = 4.616e-05 # Ablehnung
  ```
  ```

e) FLIGNER-Test für Varianzen stetiger Merkmale

- Der FLIGNER-Test oder FLIGNER-KILLEEN-Median-Test dient zum Vergleich der Varianzen mehrerer unabhängiger stetig verteilter Merkmale.
- **Geg.:** $k \geq 2$ Stichproben $x_{11}, \ldots, x_{1n_1}$ usw. bis $x_{k1}, \ldots, x_{kn_k}$ (die Stichprobenumfänge können unterschiedlich sein).
- **Vor.:** Die Zufallsgrößen X_i, $i = 1, \ldots, k$, sind unabhängig und stetig verteilt mit Varianzen σ_i^2 jeweils; repräsentative Stichproben.
- **Hyp.:** $H_0: \sigma_1^2 = \ldots = \sigma_k^2$, $H_A: \sigma_i^2 \neq \sigma_j^2$ für mindestens ein Paar (i, j).
- **R-Aufruf:** `fligner.test()` .
- Der Test ist ein rangbasierter Test, so dass Probleme bei Bindungen auftreten könnten.
- Einseitige Tests sind hier nicht möglich.

Bsp. Fligner-Test für Varianzen stetiger Merkmale

- Simulation exponentialverteilter Stichproben (unterschiedliche Varianzen) und Test auf Normalverteilung, um den stärkeren Bartlett-Test auszuschließen.
- ```
  > set.seed(123456)
  > x1=rexp(40)
  > x2=1+2*rexp(40)                        # oder x2=1+rexp(40,rate=1/2)
  > x3=2+3*rexp(50)                        # oder x3=2+rexp(40,rate=1/3)
  > shapiro.test(x1)

      Shapiro-Wilk normality test

  data: x1
  W = 0.92693, p-value = 0.0128               # <0.05/3, also Ablehnung
  ```
- Durchführung Fligner-Test, da die Voraussetzungen für den Bartlett-Test nicht erfüllt sind.
- ```
  > fligner.test(list(x1,x2,x3))

      Fligner-Killeen test of homogeneity of variances

  data: list(x1, x2, x3)
  Fligner-Killeen:med chi-squared = 23.685, df = 2, p-value = 7.191e-06
  ```

f) Zwei-Stichproben-t-Test

- Mit dem Zwei-Stichproben-t-Test wird die Gleichheit der Erwartungswerte zweier normalverteilter Merkmale mit unbekannter, aber übereinstimmender Varianz überprüft.
- **Geg.:** 2 konkrete Stichproben $x_1, \ldots, x_n$ und $y_1, \ldots, y_m$ (die Stichprobenumfänge können unterschiedlich sein).
- **Vor.:** Unabhängige normalverteilte Merkmalszufallsgrößen X und Y mit unbekannten Erwartungswerten μ_X bzw. μ_Y und unbekannter gleicher Varianz σ^2; repräsentative Stichproben.
- **Hyp.:** $H_0: \mu_X = \mu_Y$, $H_A: \mu_X \neq \mu_Y$ (zweiseitig) bzw. $H_A: \mu_X < \mu_Y$ oder $H_A: \mu_X > \mu_Y$ (einseitige Tests).
- **R-Aufruf:** `t.test(x,y,var.equal=TRUE)` bei Datenvektoren `x` und `y`.
- Ausreißer in den Daten können Probleme bereiten.

Bsp. Zwei-Stichproben-t-Test

- Simulation unabhängiger normalverteilter Stichproben mit unterschiedlichen Erwartungswerten und Test auf Normalverteilung.

-
```
> set.seed(123456)
> x=rnorm(40)                                                        # N(0,1)
> y=rnorm(50,mean=1,sd=1)                                            # N(1,1)
> shapiro.test(x)

    Shapiro-Wilk normality test

data: x
W = 0.97417, p-value = 0.4826               # >0.05/2, also Nichtablehnung

> shapiro.test(y)

    Shapiro-Wilk normality test

data: y
W = 0.98719, p-value = 0.8601               # >0.05/2, also Nichtablehnung
```

- Test auf Gleichheit der Varianzen und Zwei-Stichproben-t-Test.

Fortsetzung Bsp. Zwei-Stichproben-t-Test

-
```
> var.test(x,y)

    F test to compare two variances

data: x and y
F = 1.3586, num df = 39, denom df = 49, p-value = 0.308       # Nichtabl.
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7513357 2.5083115
sample estimates:
ratio of variances
    1.358571
```

-
```
> t.test(x,y,var.equal=TRUE)

  Two Sample t-test

data: x and y
t = -2.6898, df = 88, p-value = 0.008554                     # Ablehnung
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9881580 -0.1484186
sample estimates:
 mean of x   mean of y
0.2061461 0.7744344
```

g) Welchs-t-Test

- Mit Welchs-t-Test wird die Gleichheit der Erwartungswerte zweier normalverteilter Merkmale mit unbekannten Varianzen überprüft.
- **Geg.:** 2 konkrete Stichproben $x_1, \ldots, x_n$ und $y_1, \ldots, y_m$ (die Stichprobenumfänge können unterschiedlich sein).
- **Vor.:** Unabhängige normalverteilte Merkmalszufallsgrößen X und Y mit unbekannten Erwartungswerten μ_X bzw. μ_Y und unbekannten Varianzen σ_X^2 bzw. σ_Y^2; repräsentative Stichproben.
- **Hyp.:** $H_0: \mu_X = \mu_Y$, $H_A: \mu_X \neq \mu_Y$ (zweiseitig) bzw. $H_A: \mu_X < \mu_Y$ oder $H_A: \mu_X > \mu_Y$ (einseitige Tests).
- **R-Aufruf:** `t.test(x,y)` oder `t.test(x,y,var.equal=FALSE)` bei Datenvektoren `x` und `y`.
- Ausreißer in den Daten können Probleme bereiten.
- Der Test ist ein asymptotischer Test.

Bsp. Welchs-t-Test

- Simulation unabhängiger normalverteilter Stichproben mit unterschiedlichen Erwartungswerten und Varianzen und Test auf Normalverteilung.
-
```
> set.seed(123456)
> x=rnorm(40)                                          # N(0,1)
> y=rnorm(50,mean=1,sd=0.5)                            # N(1,0.25)
> shapiro.test(x)

    Shapiro-Wilk normality test

data: x
W = 0.97417, p-value = 0.4826          # >0.05/2, also Nichtablehnung

> shapiro.test(y)

    Shapiro-Wilk normality test

data: y
W = 0.98719, p-value = 0.8601          # >0.05/2, also Nichtablehnung
```
- Test auf Gleichheit der Varianzen und (da Ablehnung) Welchs-t-Test.

Fortsetzung Bsp. Welchs-t-Test

- ```
  > var.test(x,y)

      F test to compare two variances

  data: x and y
  F = 5.4343, num df = 39, denom df = 49, p-value = 5.776e-08    # Ablehnung
  alternative hypothesis: true ratio of variances is not equal to 1
  95 percent confidence interval:
   3.005343 10.033246
  sample estimates:
  ratio of variances
      5.434285
  ```
- ```
  > t.test(x,y)                              # oder t.test(x,y,var.equal=FALSE)

    Welch Two Sample t-test

  data: x and y
  t = -3.7294, df = 50.458, p-value = 0.0004869                   # Ablehnung
  alternative hypothesis: true difference in means is not equal to 0
  95 percent confidence interval:
   -1.0477975 -0.3143447
  sample estimates:
   mean of x   mean of y
  0.2061461 0.8872172
  ```

h) Einfache Varianzanalyse (ANOVA)

- Die einfache Varianzanalyse (ANOVA, von "analysis of variance") dient zum Test auf Gleichheit der Erwartungswerte mehrerer unabhängiger normalverteilter Merkmale.
- **Geg.:** k Stichproben $x_{11}, \dots, x_{1n_1}$ usw. bis $x_{k1}, \dots, x_{kn_k}$ (die Stichprobenumfänge können unterschiedlich sein).
- **Vor.:** Die Zufallsgrößen X_i, $i = 1, \dots, k$, sind unabhängig und normalverteilt mit Erwartungswerten μ_i jeweils und Varianz σ^2 (unbekannt, aber übereinstimmend); repräsentative Stichproben.
- **Hyp.:** $H_0: \ \mu_1 = \ldots = \mu_k$, $\quad H_A: \ \mu_i \neq \mu_j$ für mindestens ein Paar (i, j).
- **R-Aufruf:** `anova()`.
- Der $p-$Wert kann unter `Pr(>F)` abgelesen werden.
- Ausreißer in den Daten können Probleme bereiten.
- Einseitige Tests sind hier nicht möglich.

Bsp. Einfache Varianzanalyse

- Wir wenden die einfache Varianzanalyse auf die Breite des Kelchblattes ("Sepal.Width") des Iris-Beispieldatensatzes an. Dabei erhält man 3 unabhängige Stichproben, wenn man dieses Merkmal jeweils für eine der 3 untersuchten Arten beobachtet.
-
```
> data(iris)                      # Laden, dann Tests auf Normalverteilung
> shapiro.test(iris$Sepal.Width[1:50])

    Shapiro-Wilk normality test

data: iris$Sepal.Width[1:50]
W = 0.9717, p-value = 0.2715                     # >0.05/3, Nichtablehnung

> shapiro.test(iris$Sepal.Width[51:100])

    Shapiro-Wilk normality test

data: iris$Sepal.Width[51:100]
W = 0.9741, p-value = 0.338                      # >0.05/3, Nichtablehnung

> shapiro.test(iris$Sepal.Width[101:150])

    Shapiro-Wilk normality test

data: iris$Sepal.Width[101:150]
W = 0.9674, p-value = 0.1809                     # >0.05/3, Nichtablehnung
```

Fortsetzung Bsp. Einfache Varianzanalyse

- Test auf Gleichheit der Varianzen .
- ```
  > bartlett.test(Sepal.Width~Species,data=iris)

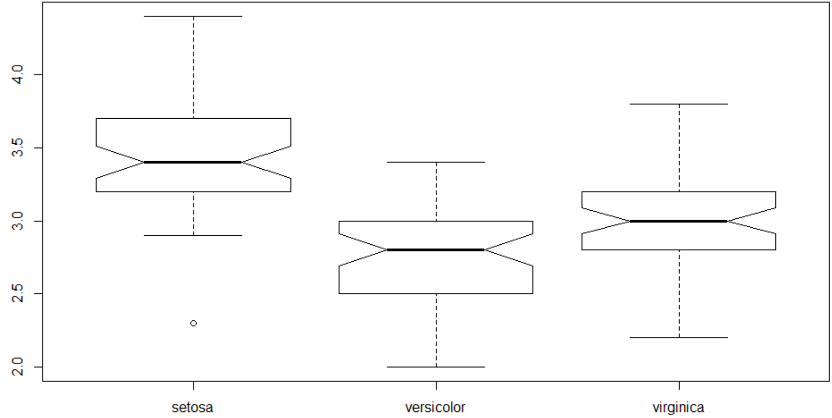
      Bartlett test of homogeneity of variances

  data: Sepal.Width by Species
  Bartlett's K-squared = 2.0911, df = 2, p-value = 0.3515   # Nichtablehnung
  ```
- ANOVA.
- ```
  > anova(lm(Sepal.Width~Species,data=iris))
  Analysis of Variance Table

  Response: Sepal.Width
              Df  Sum Sq  Mean Sq  F value     Pr(>F)
   Species     2  11.345   5.6725    49.16  < 2.2e-16  ***
   Residuals 147  16.962   0.1154
  ---
  Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  ```
- **Bemerkung:** Im `anova`-Aufruf steht `lm()` für "linear model".

Parallele Box-Plots zum Anwendungsbeispiel

```
> boxplot(Sepal.Width~Species,data=iris,notch=TRUE)
```



i) Wilcoxon-Rang-Summen-Test

- Mit dem Wilcoxon-Rang-Summen-Test vergleicht man die Lageparameter zweier Merkmale mit stetiger Verteilung miteinander.
- **Geg.:** 2 konkrete Stichproben $x_1, \ldots, x_n$ und $y_1, \ldots, y_m$ (die Stichprobenumfänge können unterschiedlich sein).
- **Vor.:** unabhängige stetig verteilte Zufallsgrößen X und Y mit Verteilungsfunktionen $F_X(x)$ und $F_Y(x) = F_X(x+c)$, $x \in \mathbb{R}$; repräsentative Stichproben.
- **Hyp.:**
  $H_0 : c = 0$, d.h. $F_X(x) = F_Y(x)$ für alle $x \in \mathbb{R}$,
  $H_A : c \neq 0$, d.h. $F_X(x) = F_Y(x-c)$ für alle $x \in \mathbb{R}$.
- **R-Aufruf:** `wilcox.test(x,y)` bei Datenvektoren `x` und `y`.
- Wird die Nullhypothese abgelehnt, kann man auf unterschiedliche Lageparameter schließen. Auch einseitige Tests sind möglich.
- Dieser Test ist ein rangbasierter Test. Bindungen können problematisch sein.

Bsp. Wilcoxon-Rang-Summen-Test

- Simulation unabhängiger exponentialverteilter Beobachtungswerte mit unterschiedlichen Erwartungswerten (Medianen,...), dann Test auf Normalverteilung.
- ```
  > set.seed(123456)
  > x=rexp(40)
  > y=1+rexp(50)
  > shapiro.test(x)

      Shapiro-Wilk normality test

  data: x
  W = 0.92693, p-value = 0.0128                    # <0.05/2, also Ablehnung
  ```
- Wilcoxon-Rang-Summen-Test.
- ```
  > wilcox.test(x,y)

      Wilcoxon rank sum test with continuity correction
  data: x and y
  W = 279, p-value = 4.903e-09                            # Ablehnung
  alternative hypothesis: true location shift is not equal to 0
  ```

j) Kruskal-Wallis-Test

- Der Kruskal-Wallis-Test dient zum Vergleich der Lage mehrerer stetiger Merkmale, er verallgemeinert den Wilcoxon-Rang-Summen-Test.
- **Geg.:** k Stichproben $x_{11}, \ldots, x_{1n_1}$ usw. bis $x_{k1}, \ldots, x_{kn_k}$.
- **Vor.:** Die Zufallsgrößen X_i , $i = 1, \ldots, k$, sind unabhängig und stetig verteilt mit Verteilungsfunktionen F_i jeweils, so dass gilt $F_i(x) = F_j(x + c_{ij})$ für alle $x \in \mathbb{R}$ mit Konstanten $c_{ij} \in \mathbb{R}$; repräsentative Stichproben.
- **Hyp.:**
  H_0 : $c_{ij} = 0$ für alle $i \neq j$,
  H_A : $c_{ij} \neq 0$ für mindestens ein Paar (i, j) .
- **R-Aufruf:** `kruskal.test()` .
- Dieser Test ist ein rangbasierter Test. Bindungen können problematisch sein.

Bsp. Kruskal-Wallis-Test

- Simulation exponentialverteilter Stichproben (unterschiedliche Varianzen) und Test auf Normalverteilung, um die ANOVA auszuschließen.
- ```
  > set.seed(123456)
  > x1=rexp(40)
  > x2=rexp(50)
  > x3=1+rexp(50)
  > shapiro.test(x1)

      Shapiro-Wilk normality test

  data: x1
  W = 0.92693, p-value = 0.0128                  # <0.05/3, also Ablehnung
  ```
- Kruskal-Wallis-Test.
- ```
  > kruskal.test(list(x1,x2,x3))

      Kruskal-Wallis rank sum test

  data: list(x1, x2, x3)
  Kruskal-Wallis chi-squared = 29.817, df = 2, p-value = 3.351e-07
  ```

5.1.4 Weitere ausgewählte statistische Tests
a) Binomialtest

- Der Binomialtest ist ein Test für die Erfolgswahrscheinlichkeit (den Parameter p) einer Bernoulli-verteilten Zufallsgröße (und damit einer diskreten Zufallsgröße).
- **Geg.:** Anzahl k der „Erfolge“ in einer konkreten Stichprobe vom Umfang n.
- **Vor.:** Die Merkmalszufallsgröße X ist Bernoulli-verteilt mit dem unbekannten Parameter p; repräsentative Stichprobe.
- **Hyp.:**
  $H_0: \; p = p_0$ (p_0 ist ein Vorgabewert für p),
  $H_A: \; p \neq p_0$ (zweiseitig) bzw.
  $H_A: \; p < p_0$ oder $H_A: \; p > p_0$ (einseitig).
- **R-Aufruf:** `binom.test(k,n,p0,...)`

Bsp. Binomialtest

- In einer Stichprobe von 100 Erzeugnissen wurden bei der Qualitätskontrolle 6 Ausschussteile gefunden. Kann man in dieser Situation von einer maximalen Sollausschussquote von 5% ausgehen oder muss man von einer größeren ausgehen?
- ```
  > binom.test(6,100,0.05,alternative="greater")

      Exact binomial test

  data: 6 and 100
  number of successes = 6, number of trials = 100, p-value = 0.384
  alternative hypothesis: true probability of success is greater than 0.05
  95 percent confidence interval:
   0.02644971 1.00000000
  sample estimates:
  probability of success
  0.06
  ```
- Keine Ablehnung (Annahme) von $H_0 : p = 0.05$, d.h. die Ausschussquote ist nicht signifikant größer als 5%.
  ```

b) Korrelations- und Abhängigkeitstests

- Zwei weitere wichtige Gruppen von Tests sind die Korrelations- und Abhängigkeitstests.
- Für mehrdimensional normalverteilte Daten kann man den PEARSON-Korrelationstest nutzen, um den Vorgabewert $\varrho_0 = 0$ für den gewöhnlichen Korrelationskoeffizienten zu überprüfen. Wird die Hypothese $H_0 : \mathbf{Corr}[X, Y] = 0$ abgelehnt, werden die normalverteilten Merkmale X und Y nicht als unabhängig angesehen.
- Für nichtnormalverteilte Zufallsvektoren kann man mit dem SPEARMAN-Korrelationstest den Vorgabewert 0 für den SPEARMANschen Rangkorrelationskoeffizienten $r^{(S)}_{X,Y}$ und damit die Unabhängigkeit der Merkmale überprüfen.
- Die Unabhängigkeit zweier kategorieller Merkmale überprüft man mit dem χ^2-Unabhängigkeitstest oder mit FISHERs exaktem Test, falls dichotome Merkmale vorliegen.

c) PEARSON-Korrelationstest

- Mit dem PEARSON-Korrelationstest überprüft man, ob der (gewöhnliche oder PEARSON-) Korrelationskoeffizient $\mathbf{Corr}[X, Y] = \rho_{(X,Y)}$ eines normalverteilten Zufallsvektors (X, Y) Null ist (dann sind die Komponenten X und Y auch stochastisch unabhängige Zufallsgrößen).
- **Geg.:** konkrete Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$.
- **Vor.:** Der Zufallsvektor (X, Y) hat eine zweidimensionale Normalverteilung mit unbekannten Parametern; repräsentative Stichprobe.
- **Hyp.:**
  $H_0: \ \mathbf{Corr}[X, Y] = 0\,, \quad H_A: \ \mathbf{Corr}[X, Y] \neq 0$ (zweiseitig) bzw.
  $H_A: \ \mathbf{Corr}[X, Y] < 0$ oder $H_A: \ \mathbf{Corr}[X, Y] > 0$ (einseitig).
- **R-Aufruf:** `cor.test(x,y)` (bei Datenvektoren `x` und `y`).
- Ausreißer in den Daten können Probleme bereiten.

Bsp. Pearson-Korrelationstest

- Simulation und Test auf Normalverteilung
- ```
  > set.seed(123456)
  > x=rnorm(50)                                  # Simulation N(0,1)
  > y=rnorm(50)                                  # Simulation N(0,1)
  > shapiro.test(x)                    # Test auf Normalverteilung X

      Shapiro-Wilk normality test

  data: x
  W = 0.97062, p-value = 0.2453             # >0.05/2, Nichtablehnung

  > shapiro.test(y)                    # Test auf Normalverteilung Y

      Shapiro-Wilk normality test

  data: y
  W = 0.97955, p-value = 0.5331             # >0.05/2, Nichtablehnung
  ```
  ```

Fortsetzung Bsp. PEARSON-Korrelationstest

- PEARSON-Korrelationstest.
-
```
> cor.test(x,y)

    Pearson's product-moment correlation

data: x and y
t = -0.20674, df = 48, p-value = 0.8371                 # Nichtablehnung
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3056375 0.2506010
sample estimates:
    cor
-0.02982726
```
- Der Zufallsvektor $(X+Y,Y)$ ist, falls X und Y unabhängige standardnormalverteilte Zufallsgrößen sind, wieder ein normalverteilter Zufallsvektor, der Korrelationskoeffizient zwischen $X+Y$ und Y ist jetzt positiv.

Fortsetzung Bsp. Pearson-Korrelationstest

- Test auf Normalverteilung von $X + Y$.
-
```
> shapiro.test(x+y)                    # Test auf Normalverteilung X+Y

    Shapiro-Wilk normality test

data: x + y
W = 0.98991, p-value = 0.9441                # >0.05/2, Nichtablehnung
```
- Einseitiger Pearson-Korrelationstest für $X + Y$ und Y.
-
```
> cor.test(x+y,y,alternative="greater")

    Pearson's product-moment correlation

data: x + y and y
t = 6.0679, df = 48, p-value = 9.873e-08               # Ablehnung
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 0.5011677 1.0000000
sample estimates:
    cor
0.6588568
```

d) Spearman-Korrelationstest

- Mit dem Spearman-Korrelationstest überprüft man, ob der Spearmansche Rangkorrelationskoeffizient $\varrho^{(S)}(X, Y)$ eines Zufallsvektors (X, Y) Null ist. Wird diese Hypothese nicht angenommen, werden die Komponenten X und Y als stochastisch abhängige Zufallsgrößen angesehen.
- **Geg.:** konkrete Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$.
- **Vor.:** Der Zufallsvektor (X, Y) hat eine zweidimensionale stetige Verteilung; repräsentative Stichprobe.
- **Hyp.:**
  $H_0: \ \varrho^{(S)}(X, Y) = 0$, $\quad H_A: \ \varrho^{(S)}(X, Y) \neq 0$ (zweiseitig) bzw.
  $H_A: \ \varrho^{(S)}(X, Y) < 0$ oder $H_A: \ \varrho^{(S)}(X, Y) > 0$ (einseitig).
- **R-Aufruf:** `cor.test(x,y,method="spearman")` (bei Datenvektoren `x` und `y`).
- Bindungen können problematisch sein.

Bsp. SPEARMAN-Korrelationstest

```
> set.seed(123456)
> x=rlnorm(50)                              # Simulation lognormale Werte
> y=rlnorm(50)                             # unabhängige lognormale Werte
> shapiro.test(x)                            # Test auf Normalverteilung

	Shapiro-Wilk normality test

data: x
W = 0.67493, p-value = 3.017e-09                     # <0.05/2, Ablehnung

> cor.test(x,y,method="spearman")

	Spearman's rank correlation rho

data: x and y
S = 20424, p-value = 0.8943                              # Nichtablehnung
alternative hypothesis: true rho is not equal to 0
sample estimates:
	rho
0.0192557
```

e) χ^2-Test auf Unabhängigkeit in Kontingenztafeln

- Der χ^2-Test auf Unabhängigkeit in Kontingenztafeln oder χ^2-Unabhängigkeitstest testet auf Unabhängigkeit zweier kategorieller Merkmale.
- **Geg.:** konkrete Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ oder Kontingenztafel, d.h. Tabelle mit den Häufigkeiten der Wertekombinationen in der Stichprobe.
- **Vor.:** kategorielle Merkmale X und Y; repräsentative Stichprobe.
- **Hyp.:**
  H_0 : X und Y sind stochastisch unabhängig;
  H_A : X und Y sind stochastisch abhängig.
- **R-Aufruf:** `chisq.test(x,y)` oder `chisq.test(table(x,y))` (bei Datenvektoren `x` und `y`).
- Der Test ist ein asymptotischer Test. Die theoretischen Häufigkeiten von Merkmalskombinationen sollten unter H_0 möglichst den Wert 5 nicht unterschreiten.

Bsp. χ^2-Test auf Unabhängigkeit in Kontingenztafeln

```
> set.seed(123456)
> x=sample(c("a","b","c"),200,TRUE,c(0.2,0.2,0.6))
> y=sample(c("u","v","w"),200,TRUE,c(0.5,0.2,0.3))
> table(x,y)
    y
 x      u   v   w
    a  23   6  13
    b  22   8   9
    c  60  26  33

> chisq.test(table(x,y))

    Pearson's Chi-squared test

data: table(x, y)
X-squared = 1.6086, df = 4, p-value = 0.8073                 # Nichtablehnung


> chisq.test(x,y)                                   # andere Form des Aufrufes

    Pearson's Chi-squared test

data: table(x, y)
X-squared = 1.6086, df = 4, p-value = 0.8073
```

f) Fishers exakter Test

- Mit Fishers exaktem Test prüft man die Unabhängigkeit zweier dichotomer Merkmale.
- **Geg.:** konkrete Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ oder $2 \times 2-$Kontingenztafel, d.h. Tabelle (mit 2 Datenzeilen und -spalten) mit den Häufigkeiten der Wertekombinationen in der Stichprobe (auch „Vierfeldertafel" genannt).
- **Vor.:** Dichotome Merkmale X und Y (nur zwei mögliche Werte jeweils); repräsentative Stichprobe.
- **Hyp.:**
  H_0 : X und Y sind stochastisch unabhängig;
  H_A : X und Y sind stochastisch abhängig.
- **R-Aufruf:** `fisher.test(x,y)` oder `fisher.test(table(x,y))` (bei Datenvektoren `x` und `y`).

Bsp. Fishers exakter Test

```
> set.seed(123456)
> x=sample(c("a","b"),40,TRUE,c(0.3,0.7))
> y=sample(c("u","v"),40,TRUE,c(0.8,0.2))
> table(x,y)
     y
 x       u   v
     a   14  5
     b   18  3
> fisher.test(x,y)

    Fisher's Exact Test for Count Data

data: x and y
p-value = 0.442                                                  # Nichtablehnung
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.06287236 2.93911606
sample estimates:
odds ratio
 0.4757131

# identische Ausgabe bei Aufruf   > fisher.test(table(x,y))
```

5.1.5 Weitere Bemerkungen zu Tests

- Statistische Tests, bei denen die Testgrößen mit Hilfe von $\overline{X}$ oder/und S^2 berechnet werden (dies sind z.B. oft Tests mit Normalverteilungsvoraussetzung), haben oft Probleme, wenn bei größerem Stichprobenumfang Ausreißer in den Daten zu finden sind. Gibt es nicht zu viele Ausreißer und liegen diese nicht weit von den Ausreißergrenzen entfernt, kann man aber häufig noch mit diesen Tests arbeiten.
- Analog verfälschen Bindungen (die z.B. durch Rundung der Realisierungen stetiger Zufallsgrößen entstehen) die Ergebnisse von rangbasierten Tests. Für eine Reihe von Tests gibt es deshalb zu berücksichtigende Korrekturterme, um trotzdem zuverlässige Resultate zu erzielen.
- Ist bei einem einseitigen Test der $p-$Wert größer als 0.5, sollte man noch einmal alles genau überprüfen, oft hat sich dann ein Fehler eingeschlichen bzw. Ausreißer verfälschen die Ergebnisse (oder das Testergebnis ist trivial).

Homoskedastizität und Heteroskedastizität

- Bei einigen statistischen Tests und Modellen wird überprüft (siehe $F-$Test, Bartlett- oder Fligner-Test) bzw. vorausgesetzt oder verlangt (siehe Zwei-Stichproben-t-Test oder ANOVA), dass die Varianzen beteiligter Zufallsgrößen übereinstimmen. Dieses nennt man auch Homoskedastizität oder Varianzhomogenität.
- Im Falle unterschiedlicher Varianzen der relevanten Zufallsgrößen spricht man dagegen auch von Heteroskedastizität oder Varianzheterogenität bzw. Varianzinhomogenität.